



<p>(51) 国際特許分類6 C12N 15/00, G06F 17/30</p>	<p>A1</p>	<p>(11) 国際公開番号 WO00/11145</p> <p>(43) 国際公開日 2000年3月2日(02.03.00)</p>
<p>(21) 国際出願番号 PCT/JP99/01693</p> <p>(22) 国際出願日 1999年3月31日(31.03.99)</p> <p>(30) 優先権データ 特願平10/232901 1998年8月19日(19.08.98) JP</p> <p>(71) 出願人 (米国を除くすべての指定国について) 富士通株式会社(FUJITSU LIMITED)[JP/JP] 〒211-8588 神奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa, (JP)</p> <p>(72) 発明者; および (75) 発明者/出願人 (米国についてのみ) 川西祐一(KAWANISHI, Yuichi)[JP/JP] 内藤公敏(NAITO, Kimitoshi)[JP/JP] 宮澤 慎(MIYAZAWA, Shin)[JP/JP] 伊藤辰也(ITO, Tatsuya)[JP/JP] 二木敬正(FUTATSUKI, Takamasa)[JP/JP] 〒211-8588 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内 Kanagawa, (JP)</p>		<p>(74) 代理人 笹島富二雄(SASAJIMA, Fujio) 〒105-0001 東京都港区虎ノ門1丁目19番5号 虎ノ門1丁目森ビル Tokyo, (JP)</p> <p>(81) 指定国 US, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE)</p> <p>添付公開書類 国際調査報告書</p>
<p>(54) Title: APPARATUS FOR EXTRACTING AND PROCESSING GENE MOTIF, METHOD FOR EXTRACTING AND PROCESSING GENE MOTIF AND RECORDING MEDIUM CARRYING GENE MOTIF EXTRACTION/PROCESSING PROGRAM</p> <p>(54) 発明の名称 遺伝子のモチーフ抽出処理装置及び遺伝子のモチーフ抽出処理方法並びに遺伝子のモチーフ抽出処理プログラムを記録した記録媒体</p> <p>(57) Abstract Based on at least one of gene sequence data having been input, a gene motif is extracted. Based on the thus extracted motif, gene sequence data containing this motif are detected from gene sequence data having been clarified. Next, the thus detected gene sequence data are added to the gene sequence data having been input, thereby elevating the gene motif extraction efficiency.</p> <div data-bbox="760 1276 1396 1659"> </div> <p>10 ... DATA INPUTTING UNIT 12 ... DATA EDITING UNIT 14 ... AUTOMATIC ALIGNMENT UNIT 16 ... MOTIF EXTRACTING UNIT 18 ... EXTRACTION DATA PROCESSING UNIT 20 ... MOTIF DETECTING UNIT 26 ... MOTIF DB 28a ... GENE SEQUENCE DB 28b ... STEREOISOMER DB</p>		

(57)要約

入力された少なくとも1つの遺伝子配列情報から遺伝子のモチーフを抽出し、抽出されたモチーフに基づいて、モチーフを一部に含む遺伝子配列情報を、解明済みの遺伝子配列情報から検索する。そして、検索された遺伝子配列情報を、入力された遺伝子配列情報に付け加えることで、遺伝子のモチーフ抽出効率を向上させる。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE	アラブ首長国連邦	DM	ドミニカ	KZ	カザフスタン	RU	ロシア
AL	アルバニア	EE	エストニア	LC	セントルシア	SD	スーダン
AM	アルメニア	ES	スペイン	LI	リヒテンシュタイン	SE	スウェーデン
AT	オーストリア	FI	フィンランド	LK	スリ・ランカ	SG	シンガポール
AU	オーストラリア	FR	フランス	LR	リベリア	SI	スロヴェニア
AZ	アゼルバイジャン	GA	ガボン	LS	レソト	SK	スロヴァキア
BA	ボスニア・ヘルツェゴビナ	GB	英国	LT	リトアニア	SL	シエラ・レオネ
BB	バルバドス	GD	グレナダ	LU	ルクセンブルグ	SN	セネガル
BE	ベルギー	GE	グルジア	LV	ラトヴィア	SZ	スワジランド
BF	ブルキナ・ファソ	GH	ガーナ	MA	モロッコ	TD	チャード
BG	ブルガリア	GM	ガンビア	MC	モナコ	TG	トーゴ
BJ	ベナン	GN	ギニア	MD	モルドヴァ	TJ	タジキスタン
BR	ブラジル	GW	ギニア・ビサウ	MG	マダガスカル	TZ	タンザニア
BY	ベラルーシ	GR	ギリシャ	MK	マケドニア旧ユーゴスラヴィア共和国	TM	トルクメニスタン
CA	カナダ	HR	クロアチア	ML	マリ	TR	トルコ
CF	中央アフリカ	HU	ハンガリー	MN	モンゴル	TT	トリニダード・トバゴ
CG	コンゴ	ID	インドネシア	MR	モーリタニア	UA	ウクライナ
CH	スイス	IE	アイルランド	MW	マラウイ	UG	ウガンダ
CI	コートジボワール	IL	イスラエル	MX	メキシコ	US	米国
CM	カメルーン	IN	インド	NE	ニジェール	UZ	ウズベキスタン
CN	中国	IS	アイスランド	NL	オランダ	VN	ヴェトナム
CR	コスタ・リカ	IT	イタリア	NO	ノルウェー	YU	ユーゴスラビア
CU	キューバ	JP	日本	NZ	ニュージーランド	ZA	南アフリカ共和国
CY	キプロス	KE	ケニア	PL	ポーランド	ZW	ジンバブエ
CZ	チェコ	KG	キルギスタン	PT	ポルトガル		
DE	ドイツ	KR	北朝鮮	RO	ルーマニア		
DK	デンマーク		韓国				

## 明 細 書

遺伝子のモチーフ抽出処理装置及び遺伝子のモチーフ抽出処理方法並びに遺伝子のモチーフ抽出処理プログラムを記録した記録媒体

### 技術分野

本発明は、遺伝子のモチーフを自動的に抽出する技術に関し、特に、遺伝子のモチーフ抽出効率を向上させる技術に関する。

### 背景技術

近年の遺伝子工学の進歩に伴い、DNA配列やアミノ酸配列で表現される遺伝子配列の決定技術が急速に進歩している。また、人類を初めとする様々な生物種を対象として、特定生物の遺伝子配列を全て解明しようとするゲノムプロジェクトが世界規模で行われている。このため、解明済みの遺伝子配列情報を有効に活用すべく、遺伝子配列情報データベースが爆発的な勢いで増加している。

これらの遺伝子配列の中には、配列情報は明らかになっているが、その機能や構造に関しては未知であるものも多い。かかる遺伝子の機能や構造を配列から推測する有効な方法として、特徴的な規則性を持つモチーフの抽出がある。このため、本願出願人は先の出願において、複数の遺伝子配列情報を比較することで、モチーフを自動的に抽出する技術を提案した（特開平7-274965号公報参照）。

しかしながら、かかるモチーフ抽出技術では、抽出されたモチーフは、プリンタから出力されるか、或いは、データベース等へファイルとして出力されるだけであって、再利用が充分なされていないのが現状であった。このため、遺伝子の機能や構造の解明を進めるべく、抽出されたモチーフを遺伝子配列情報に加えて、再度モチーフの抽出を行おうとする場合には、抽出されたモチーフ情報を作業者が入力しなければならなかった。従って、従来の技術では、モチーフ抽出効率の向上には自ずから限界があり、さらなるモチーフ抽出効率の向上は困難であった。

## 2

そこで、本発明は以上のような従来の問題点に鑑み、抽出されたモチーフを再利用する機構を付加することにより、遺伝子の機能や構造を推測するためのモチーフ抽出効率を向上させる技術を提供することを目的とする。

## 発明の開示

上記目的達成のため、第1の手段として、解明済みの遺伝子配列情報を蓄積する遺伝子配列情報蓄積手段と、少なくとも1つの遺伝子配列情報を入力する遺伝子配列情報入力手段と、該遺伝子配列情報入力手段により入力された遺伝子配列情報から、遺伝子のモチーフを抽出するモチーフ抽出手段と、該モチーフ抽出手段により抽出されたモチーフに基づいて、該モチーフを一部に含む遺伝子配列情報を、前記遺伝子配列情報蓄積手段から検索する遺伝子配列情報検索手段と、該遺伝子配列情報検索手段により検索された遺伝子配列情報を、前記遺伝子配列情報入力手段により入力された遺伝子配列情報に付け加える遺伝子配列情報付加手段と、を含んで遺伝子のモチーフ抽出処理装置を構成したことを特徴とする。

かかる構成によれば、遺伝子のモチーフを抽出する場合には、先ず、遺伝子配列情報入力手段を介して、少なくとも1つの遺伝子配列情報を入力する。すると、モチーフ抽出手段により、入力された遺伝子配列情報から遺伝子のモチーフが抽出される。そして、遺伝子配列情報検索手段により、抽出されたモチーフを一部に含む遺伝子配列情報が遺伝子配列情報蓄積手段から検索される。検索された遺伝子配列情報は、必要に応じて、遺伝子配列情報付加手段により、遺伝子配列情報入力手段により入力された遺伝子配列情報に付け加えられる。その後、モチーフ抽出手段によるモチーフの抽出、遺伝子配列情報検索手段による遺伝子配列情報の検索を繰り返すことによって、遺伝子配列情報の機能や構造を次第に解明していく。即ち、抽出されたモチーフを一部に含む遺伝子配列情報は、入力された遺伝子配列情報に付け加えられるので、抽出されたモチーフの再利用が行われるようになり、遺伝子の機能や構造を推測するためのモチーフ抽出効率を向上させることができる。

また、前記遺伝子配列情報入力手段により入力された遺伝子配列情報の中から、

## 3

前記モチーフ抽出手段によるモチーフの抽出範囲を指定するモチーフ抽出範囲指定手段を備え、前記モチーフ抽出手段は、前記モチーフ抽出範囲指定手段により指定された抽出範囲内で、遺伝子のモチーフを抽出する構成としてもよい。

かかる構成によれば、モチーフ抽出範囲指定手段によりモチーフの抽出範囲を指定すると、指定された抽出範囲内で、遺伝子のモチーフが抽出される。従って、例えば、入力された種々の生物の遺伝子配列情報の中から、類似する生物の遺伝子配列情報のみを抽出範囲とすることができ、遺伝子配列の機能や構造の解明をより容易に行なうことができる。

さらに、前記遺伝子配列情報を編集する遺伝子配列情報編集手段を備えた構成としてもよい。

かかる構成によれば、遺伝子配列情報に対して、遺伝子配列情報編集手段により各種編集が行われるので、利用者の意思に沿ったモチーフの抽出を行うことができる。

この他には、前記モチーフ抽出手段により抽出されたモチーフを編集するモチーフ編集手段を備えた構成としてもよい。

かかる構成によれば、抽出されたモチーフに対して、モチーフ編集手段により各種編集が行われるので、利用者の意思に沿った遺伝子配列情報の検索を行うことができる。

また、前記遺伝子配列情報入力手段により入力された複数の遺伝子配列情報を、アラインメント処理するアラインメント手段を備えた構成としてもよい。

かかる構成によれば、入力された複数の遺伝子配列情報は、アラインメント手段によりアラインメント処理されるので、必要と思われる遺伝子配列情報を順不同で入力すればよく、遺伝子配列情報の入力作業の効率を向上することができる。また、利用者は、入力された複数の遺伝子配列情報間で類似している領域を視覚的に把握できるため、前記モチーフ抽出範囲指定手段によるモチーフの抽出範囲の指定を容易に行うことができる。

さらに、モチーフを蓄積するモチーフ蓄積手段と、前記モチーフ抽出手段により抽出されたモチーフを、該モチーフ蓄積手段に登録するモチーフ登録手段と、

を備えた構成としてもよい。

かかる構成によれば、抽出されたモチーフは、モチーフ登録手段によりモチーフ蓄積手段に登録されるため、他の作業者が、抽出されたモチーフを再利用することが可能となり、遺伝子配列情報の機能や構造を解明する作業の効率をより向上することができる。

ここで、前記モチーフ蓄積手段に登録されているモチーフの中から、少なくとも1つのモチーフを表示するモチーフ表示手段を備えた構成としてもよい。

かかる構成によれば、必要に応じて、モチーフ登録手段に登録されているモチーフが表示されるので、種々のモチーフを容易に比較することができ、遺伝子配列の機能や構造の解明をより容易に行うことができる。

第2の解決手段として、少なくとも1つの遺伝子配列情報を入力する遺伝子配列情報入力工程と、該遺伝子配列情報入力工程により入力された遺伝子配列情報から、遺伝子のモチーフを抽出するモチーフ抽出工程と、該モチーフ抽出工程により抽出されたモチーフに基づいて、該モチーフを一部に含む遺伝子配列情報を、遺伝子配列情報データベースから検索する遺伝子配列情報検索工程と、該遺伝子配列情報検索工程により検索された遺伝子配列情報を、前記遺伝子配列情報入力工程により入力された遺伝子配列情報に付け加える遺伝子配列情報付加工程と、を含んで遺伝子のモチーフ抽出処理方法を構成したことを特徴とする。

かかる構成によれば、遺伝子のモチーフを抽出する場合には、先ず、遺伝子配列情報入力工程を介して、少なくとも1つの遺伝子配列情報を入力する。すると、モチーフ抽出工程により、入力された遺伝子配列情報から遺伝子のモチーフが抽出される。そして、遺伝子配列情報検索工程により、抽出されたモチーフを一部に含む遺伝子配列情報が遺伝子配列情報データベースから検索される。検索された遺伝子配列情報は、必要に応じて、遺伝子配列情報付加工程により、遺伝子配列情報入力工程により入力された遺伝子配列情報に付け加えられる。その後、モチーフ抽出工程によるモチーフの抽出、遺伝子配列情報検索工程による遺伝子配列情報の検索を繰り返すことによって、遺伝子配列情報の機能や構造を次第に推測していく。即ち、抽出されたモチーフを一部に含む遺伝子配列情報は、入力さ

れた遺伝子配列情報に付け加えられるので、抽出されたモチーフの再利用が行われるようになり、遺伝子の機能や構造を推測するためのモチーフ抽出効率を向上させることができる。

第3の解決手段として、少なくとも1つの遺伝子配列情報を入力する遺伝子配列情報入力機能と、該遺伝子配列情報入力機能により入力された遺伝子配列情報から、遺伝子のモチーフを抽出するモチーフ抽出機能と、該モチーフ抽出機能により抽出されたモチーフに基づいて、該モチーフを一部に含む遺伝子配列情報を、遺伝子配列情報データベースから検索する遺伝子配列情報検索機能と、該遺伝子配列情報検索機能により検索された遺伝子配列情報を、前記遺伝子配列情報入力機能により入力された遺伝子配列情報に付け加える遺伝子配列情報付加機能と、を実現するための遺伝子のモチーフ抽出処理プログラムを記録媒体に記録した。

ここで、「記録媒体」とは、電子情報を確実に記録でき、かつ、必要に応じて確実に取り出し可能なものをいい、例えば、磁気テープ、磁気ディスク、磁気ドラム、ICカード、CD-ROM等の可搬記録媒体が該当する。

かかる構成によれば、遺伝子配列情報入力機能と、モチーフ抽出機能と、遺伝子配列情報検索機能と、遺伝子配列情報付加機能と、を実現するための遺伝子のモチーフ抽出処理プログラムが記録媒体に記録される。従って、かかるプログラムを記録した記録媒体があれば、一般的な電子計算機を用いて、本発明に係る遺伝子のモチーフ抽出処理装置が容易に構築される。

#### 図面の簡単な説明

図1は、本発明に係る遺伝子のモチーフ抽出処理装置の構成図である。

図2は、モチーフ抽出を行うための作業手順図である。

図3は、初期画面の説明図である。

図4は、マルチプルアラインメントデータを入力した初期画面の説明図である。

図5は、編集画面の説明図である。

図6は、モチーフ抽出範囲を配列で指定した様子の説明図である。

図7は、モチーフ抽出範囲を矩形領域で指定した様子の説明図である。

図8は、抽出結果表示画面の説明図である。

図9は、検索対象を指定した抽出結果表示画面の説明図である。

図10は、検索結果表示画面の説明図である。

図11は、再利用する配列を指定した検索結果表示画面の説明図である。

図12は、再利用する配列がマルチプルアラインメントデータに追加された様子の説明図である。

#### 発明を実施するための最良の形態

本発明をより詳細に説述するために、添付の図面に従ってこれを説明する。

図1は、本発明に係る遺伝子のモチーフ抽出処理装置（以下「モチーフ抽出処理装置」という）の構成を示す。

モチーフ抽出処理装置は、少なくとも、中央処理装置（CPU）とメモリとを備えた電子計算機から構成され、メモリ上にロードされたプログラムに従ってモチーフ抽出に伴う各種処理を実行する。モチーフ抽出処理装置は、データ入力部10（遺伝子配列情報入力手段、工程及び機能）と、データ編集部12と、自動アラインメント部14と、モチーフ抽出部16と、抽出結果処理部18と、モチーフ検索部20と、表示装置22と、印字装置24と、モチーフデータベース26（モチーフ蓄積手段）と、遺伝子配列情報データベース28（遺伝子配列情報蓄積手段）と、を含んで構成される。

データ入力部10は、例えば、キーボードとマウスとから構成され、モチーフ抽出対象となる少なくとも1つ以上の遺伝子配列データを入力するデータ入力機能を提供する。

データ編集部12は、データ入力部10を介して入力された遺伝子配列データに対して、データ編集機能（遺伝子配列情報編集手段）、抽出範囲指定機能（モチーフ抽出範囲指定手段）、パラメータ設定機能、データ表示機能、外部データ入力機能及びデータ読込機能を提供する。データ編集機能では、遺伝子配列データに対して各種編集を行うべく、例えば、入力された複数の遺伝子配列データに新たな遺伝子配列データを追加したり、入力された複数の遺伝子配列データから



## 7

任意の遺伝子配列データを削除したり、任意の遺伝子情報を置き換えたり、手作業によるアラインメント実施機能が提供される。抽出範囲指定機能では、モチーフ抽出範囲を指定すべく、複数の遺伝子配列データからモチーフ抽出対象とする遺伝子配列データを特定したり、アラインメントされた遺伝子情報の範囲を矩形で指定する機能が提供される。パラメータ設定機能では、モチーフの自動抽出を行うための各種パラメータを設定する機能が提供される。データ表示機能では、モチーフ抽出処理装置と利用者とのインターフェースを確立すべく、編集中の遺伝子配列データを表示装置 22 に表示する機能が提供される。外部データ入力機能では、検索された遺伝子配列の再利用を可能にすべく、モチーフ検索部 20 からのデータを入力する機能が提供される。データ読込機能では、既存のマルチプルアラインメントデータの有効利用を図るべく、ファイルからマルチプルアラインメントデータを読み込む機能が提供される。

自動アラインメント部 14 は、データ編集部 12 と協働して、遺伝子配列データを自動アラインメントする機能（アラインメント手段）を提供する。従って、データ編集部 12 に、アラインメント処理済みの遺伝子配列データを入力する必要はなく、例えば、ファイルから遺伝子配列データを複数読み込んだ後、自動アラインメント機能を用いてアラインメント処理を行ってもよい。

モチーフ抽出部 16 は、データ編集部 12 の抽出範囲指定機能により指定されたモチーフ抽出範囲に対して、モチーフを自動抽出する自動抽出機能（モチーフ抽出手段、工程及び機能）を提供する。

抽出結果処理部 18 は、モチーフ抽出部 16 により自動抽出されたモチーフ（以下「抽出モチーフ」という）に対して、データベース登録機能（モチーフ登録手段）、データベース参照機能、データ表示機能及びデータ印字機能を提供する。データベース登録機能では、抽出モチーフをモチーフデータベース 26 に登録する機能が提供される。データベース参照機能では、モチーフデータベース 26 に登録されているモチーフの中から、任意のモチーフを参照する機能が提供される。データ表示機能では、抽出モチーフ或いは参照したモチーフ（以下「参照モチーフ」という）を表示装置 22 に表示する機能が提供される。なお、データベース

参照機能及びデータ表示機能により、モチーフ表示手段が構成される。データ印字機能では、抽出モチーフ或いは参照モチーフを印字装置 24 に出力する機能が提供される。また、抽出結果処理部 18 は、データ編集部 12 のデータ編集機能を利用して、抽出モチーフを編集することができるようになっている（モチーフ編集手段）。

モチーフ検索部 20 は、遺伝子配列情報データベース 28 から抽出モチーフをその一部に含む遺伝子情報を検索すべく、検索範囲指定機能、検索実行機能（遺伝子配列情報検索手段、工程及び機能）、出力データ指定機能及びデータ表示機能を提供する。検索範囲指定機能では、複数の抽出モチーフの中から、検索を行うモチーフを指定する機能が提供される。検索実行機能では、検索範囲指定機能により指定されたモチーフに基づいて、遺伝子配列情報データベース 28 を検索する機能が提供される。出力データ指定機能では、データ編集部 12 の外部データ入力機能を介して、マルチプルアラインメントデータに追加する検索結果を指定する機能が提供される。なお、データ編集部 12 の外部データ入力機能及びモチーフ検索部 20 の出力データ指定機能により、遺伝子配列情報付加手段、工程及び機能が構成される。データ表示機能では、利用者に対するインターフェースの提供、或いは、検索結果の表示等を行うべく、表示装置 22 にデータを表示する機能が提供される。

遺伝子配列情報データベース 28 は、遺伝子配列データベース 28 a と、立体構造データベース 28 b と、を含んで構成される。遺伝子配列データベース 28 a には、解明済みの遺伝子配列情報が登録される。立体構造データベース 28 b には、解明済みの遺伝子配列情報であって、その構造を立体的に記述した座標データからなる構造情報が登録される。

次に、かかる構成からなるモチーフ抽出処理装置の作用について、図 2 の作業手順図に基づいて説明する。

ステップ 1（図では「S1」と略記する。以下同様）では、遺伝子配列データの入力が行われる。

モチーフ抽出処理装置を起動すると、図 3 に示すような初期画面 30 が表示さ

れる。初期画面 30 は、データ表示部 32 と、「ファイル入力」ボタン 34 と、「自動アラインメント」ボタン 36 と、「モチーフ抽出」ボタン 38 と、「データ編集」ボタン 40 と、「設定」ボタン 42 と、を含んで構成される。データ表示部 32 には、入力された遺伝子配列データが表示される。「ファイル入力」ボタン 34 は、ファイルから遺伝子配列データを読み込むべく、データ読込機能を起動するときにクリックされる。「自動アラインメント」ボタン 36 は、入力された遺伝子配列データに対してアラインメント処理を行うべく、自動アラインメント部 14 のアライメント機能を起動するときにクリックされる。「モチーフ抽出」ボタン 38 は、入力された遺伝子配列データからモチーフを自動抽出すべく、モチーフ抽出部 16 の自動抽出機能を起動するときにクリックされる。「データ編集」ボタン 40 は、入力された遺伝子配列データの編集を行うべく、データ編集機能を起動するときにクリックされる。「設定」ボタン 42 は、モチーフを抽出する際に必要な各種パラメータの設定を行うべく、パラメータ設定機能を起動するときにクリックされる。

遺伝子配列データを入力するには、次のような 2 通りの方法がある。第 1 の方法は、ファイルから遺伝子配列データを入力する方法である。即ち、初期画面 30 の「ファイル入力」ボタン 34 をクリックすると、図示しないファイル指定画面が表示される。そして、ファイル指定画面でファイル名を指定すると、指定したファイルから遺伝子配列データが読み込まれ、図 4 に示すように、データ表示部 32 に読み込まれた遺伝子配列データが表示される。また、第 2 の方法は、データ入力部 10 のデータ入力機能を使用して、遺伝子配列データを直接入力する方法である。即ち、初期画面 30 の「データ編集」ボタン 40 をクリックすると、図 5 に示すように、遺伝子配列データの各種編集を行う編集画面 50 が表示される（編集画面 50 の詳細は後述する）。そして、データ編集機能の一部を使用して、データ入力部 10 から遺伝子配列データを直接入力する。

ステップ 2 では、入力された遺伝子配列データの各種編集が行われる。

初期画面 30 の「データ編集」ボタン 40 がクリックされると、データ編集部 12 のデータ編集機能が起動され、図 5 に示す編集画面 50 が表示される。編集

画面50は、データ表示部52と、「データ追加」ボタン54と、「自動アラインメント」ボタン56と、「モチーフ抽出」ボタン58と、「設定」ボタン60と、を含んで構成される。ここで、データ表示部52、「自動アラインメント」ボタン56、「モチーフ抽出」ボタン58及び「設定」ボタン60は、夫々、初期画面30におけるデータ表示部32、「自動アラインメント」ボタン36、「モチーフ抽出」ボタン38及び「設定」ボタン42と実質的に同一であるので、その説明は省略する。「データ追加」ボタン54は、データ表示部52に表示されている遺伝子配列データに、ファイルから新規に遺伝子配列データを追加すべく、データ読込機能を起動するときにクリックされる。

編集画面50上では、例えば、任意の位置にギャップ（図では「-」で示す）を挿入したり、或いは、任意の遺伝子情報を置き換えることができる。また、マルチプルアラインメントデータに新たな遺伝子配列データを付け加える場合には、「データ追加」ボタン54をクリックすると、図示しないファイル指定画面が表示される。そして、ファイル指定画面でファイル名を指定すると、指定したファイルから遺伝子配列データが読み込まれ、データ表示部52に読み込まれた遺伝子配列データが表示される。なお、追加される遺伝子配列データは、マルチプルアラインメントデータであってもよい。

ステップ3では、必要に応じて、モチーフ抽出を行うマルチプルアラインメントデータの範囲指定が行われる。

範囲指定を行わない場合には、データ表示部52に表示されたマルチプルアラインメントデータ全体がモチーフ抽出対象となるが、実際には、マルチプルアラインメントデータの一部のみをモチーフ抽出対象としたい場合がある。この場合には、抽出範囲指定機能によりモチーフ抽出対象を限定すべく、次のような2通りの方法で範囲指定を行うことができる。第1の方法は、図6に示すように、マウス等のポインティングデバイスを使用して、マルチプルアラインメントデータの中から少なくとも1つのアラインメントデータを指定する方法である。また、第2の方法は、図7に示すように、ポインティングデバイスを使用して、モチーフ抽出対象を矩形領域で指定する方法である。なお、モチーフ抽出対象を変更す

る場合には、モチーフ抽出対象を指定し直せばよい。

ステップ4では、モチーフの自動抽出が行われる。

初期画面30の「モチーフ抽出」ボタン38、或いは、編集画面50の「モチーフ抽出」ボタン58がクリックされると、モチーフ抽出部16の自動抽出機能が起動され、マルチプルアラインメントデータからモチーフが自動的に抽出される。このとき、モチーフ抽出対象が指定されていれば、指定された範囲内からモチーフが抽出される。

ステップ5では、モチーフの抽出結果が表示される。

自動的に抽出されたモチーフは、抽出結果処理部18のデータ表示機能により、図8に示すような抽出結果表示画面70に表示される。抽出結果表示画面70は、抽出結果表示部72と、抽出モチーフ表示部74と、「ファイル出力」ボタン76と、「データベース検索」ボタン78と、「アラインメントデータ入力」ボタン80と、を含んで構成される。抽出結果表示部72には、抽出パラメータ72a、解析結果72b等が表示される。抽出モチーフ表示部74には、抽出されたモチーフの一覧が表示される。「ファイル出力」ボタン76は、抽出結果を印字装置24に出力すべくデータ印字機能を起動したり、或いは、抽出結果をモチーフデータベース26に登録すべくデータベース登録機能を起動するときに、クリックされる。「データベース検索」ボタン78は、遺伝子配列情報データベース28に対して検索を行うべく、モチーフ検索部20の検索実行機能を起動するときにクリックされる。「アラインメントデータ入力」ボタン80は、現在表示されているアラインメントデータを破棄し、モチーフ抽出を行うアラインメントデータをファイルから読み込む場合にクリックされる。

ステップ6では、モチーフ抽出結果の出力が行われる。

即ち、抽出結果表示画面70上に表示されているモチーフ抽出結果を、印字装置24に出力したり、或いは、モチーフデータベース26に登録する処理が行われる。モチーフ抽出結果を出力するには、抽出結果表示画面70の「ファイル出力」ボタン76をクリックすると、出力先を指定する出力先指定画面（図示せず）が表示される。そして、出力先指定画面上で出力先を指定することで、印字装

## 12

置24へのモチーフ抽出結果の出力、或いは、モチーフデータベース26へのモチーフ抽出結果の登録が行われる。

ステップ7では、既存のモチーフ抽出結果との比較が行われる。

即ち、モチーフ抽出結果と、モチーフデータベース26に既に登録されているモチーフ抽出結果と、を比較することにより、モチーフにより推測される遺伝子の機能等の特定作業が行われる。

ステップ8では、遺伝子配列情報データベース28を検索する検索対象の指定が行われる。

即ち、図9に示すように、抽出モチーフ表示部74に表示されるモチーフ一覧の中から1つのモチーフをクリックすると、その先頭部に指定されたことを示す記号（図では「\*」で表わされる）74aが付される。検索対象たるモチーフを変更したい場合には、新たなモチーフをクリックすると、そのモチーフが検索対象となる。なお、複数のモチーフを検索対象とするようにしてもよい。

ステップ9では、モチーフ検索が行われる。

即ち、検索対象たるモチーフを指定した後で、「データベース検索」ボタン78をクリックすると、指定されたモチーフに基づいて、遺伝子配列情報データベース28が検索される。このとき、遺伝子配列情報データベース28を構成する遺伝子配列データベース28a或いは立体構造データベース28bのどちらを検索するかは、「アラインメントデータ入力」ボタン80をクリックすることで任意に選択できる。

ステップ10では、遺伝子配列情報データベース28の検索結果が表示される。

遺伝子配列情報データベース28の検索結果は、モチーフ検索部20のデータ表示機能により、図10に示すような検索結果表示画面90に表示される。検索結果表示画面90は、モチーフ表示部92と、検索結果表示部94と、「ファイル出力」ボタン96と、「アラインメントへの追加」ボタン98と、「エントリ内容表示」ボタン100と、「データベース検索」ボタン102と、を含んで構成される。モチーフ表示部92には、遺伝子配列情報データベース28の検索対象たるモチーフが表示される。検索結果表示部94には、遺伝子配列情報データ

## 13

ベース28の検索結果が表示される。「ファイル出力」ボタン96は、検索結果を印字装置24に出力したり、或いは、モチーフデータベース26に登録するときにクリックされる。「アラインメントへの追加」ボタン98は、データ編集部12の外部データ入力機能を介して、マルチプルアラインメントデータに検索結果を追加するときにクリックされる。「エントリ内容表示」ボタン100は、エントリを表示するときにクリックされる。「データベース検索」ボタン102は、遺伝子配列情報データベース28に対して、絞り込み検索を行うときにクリックされる。

ステップ11では、再利用する遺伝子配列データが指定される。

即ち、検索結果表示画面90において、図11に示すように、再利用する遺伝子配列データをクリックすると、指定された遺伝子配列データの先頭部に記号「X」が付される。そして、「アラインメントへの追加」ボタン98をクリックすると、データ編集部12の外部データ入力機能を介して遺伝子配列データが取り込まれ、図12に示すように、マルチプルアラインメントデータの最下段に追加される。

その後、ステップ2～ステップ11を繰り返し、遺伝子の機能や構造を解明する。

かかるモチーフ抽出処理装置によれば、マルチプルアラインメントデータの入力から、モチーフ抽出、抽出モチーフに基づくデータベース検索といった一連の処理を統合化されたシステム上で行うことが可能となる。利用者は、データ編集部12のデータ編集機能を利用することで、入力されたマルチプルアラインメントデータに対して、モチーフ抽出を行うための詳細な設定を行うことが可能となる。また、表示装置22が提供するユーザインターフェースを利用することにより、視覚的に抽出モチーフを確認しながら再度モチーフ抽出を行ったり、或いは、以前に実行した抽出結果との比較を表示装置22上で行うことが可能となる。さらに、モチーフ検索部20の検索実行機能を利用すれば、モチーフ単位でのデータベース検索を容易に行うことができ、また、データ表示機能を利用すれば、検索されたデータのデータベースの登録内容の閲覧が簡単にできる。この結果、遺

伝子配列の機能や構造を推測するためのモチーフの抽出を効率良く、かつ、利用者の要求に応じて柔軟的に行うことが可能となる。

このような機能を実現するプログラムを、例えば、磁気テープ、磁気ディスク、磁気ドラム、ICカード、CD-ROM等の可搬記録媒体に記録しておけば、本発明に係るモチーフ抽出処理プログラムを市場に流通させることができる。そして、かかる媒体を取得した者は、一般的な電子計算機システムを利用して、モチーフ抽出処理装置を容易に構築することが可能となる。

#### 産業上の利用可能性

以上説明したように、本発明に係る遺伝子のモチーフ抽出処理装置及び遺伝子のモチーフ抽出処理方法は、遺伝子の機能や構造を推測するためのモチーフ抽出効率を向上させることができ、極めて有用なものである。また、遺伝子のモチーフ抽出処理プログラムを記録した記録媒体は、一般的な電子計算機システムを利用して遺伝子のモチーフ抽出処理装置を容易に構築できるので、極めて有用なものである。



## 請求の範囲

1. 解明済みの遺伝子配列情報を蓄積する遺伝子配列情報蓄積手段と、  
少なくとも1つの遺伝子配列情報を入力する遺伝子配列情報入力手段と、  
該遺伝子配列情報入力手段により入力された遺伝子配列情報から、遺伝子のモチーフを抽出するモチーフ抽出手段と、  
該モチーフ抽出手段により抽出されたモチーフに基づいて、該モチーフを一部に含む遺伝子配列情報を、前記遺伝子配列情報蓄積手段から検索する遺伝子配列情報検索手段と、  
該遺伝子配列情報検索手段により検索された遺伝子配列情報を、前記遺伝子配列情報入力手段により入力された遺伝子配列情報に付け加える遺伝子配列情報付加手段と、  
を含んで構成されたことを特徴とする遺伝子のモチーフ抽出処理装置。
2. 前記遺伝子配列情報入力手段により入力された遺伝子配列情報の中から、前記モチーフ抽出手段によるモチーフの抽出範囲を指定するモチーフ抽出範囲指定手段を備え、  
前記モチーフ抽出手段は、前記モチーフ抽出範囲指定手段により指定された抽出範囲内で、遺伝子のモチーフを抽出する構成である請求の範囲第1項記載の遺伝子のモチーフ抽出処理装置。
3. 前記遺伝子配列情報を編集する遺伝子配列情報編集手段を備えた構成である請求の範囲第1項記載の遺伝子のモチーフ抽出処理装置。
4. 前記モチーフ抽出手段により抽出されたモチーフを編集するモチーフ編集手段を備えた構成である請求の範囲第1項記載の遺伝子のモチーフ抽出処理装置。
5. 前記遺伝子配列情報入力手段により入力された複数の遺伝子配列情報を、アラインメント処理するアラインメント手段を備えた構成である請求の範囲第1項記載の遺伝子のモチーフ抽出処理装置。
6. モチーフを蓄積するモチーフ蓄積手段と、  
前記モチーフ抽出手段により抽出されたモチーフを、該モチーフ蓄積手段に登

録するモチーフ登録手段と、

を備えた構成である請求の範囲第1項記載の遺伝子のモチーフ抽出処理装置。

7. 前記モチーフ蓄積手段に登録されているモチーフの中から、少なくとも1つのモチーフを表示するモチーフ表示手段を備えた構成である請求の範囲第6項記載の遺伝子のモチーフ抽出処理装置。

8. 少なくとも1つの遺伝子配列情報を入力する遺伝子配列情報入力工程と、  
該遺伝子配列情報入力工程により入力された遺伝子配列情報から、遺伝子のモチーフを抽出するモチーフ抽出工程と、

該モチーフ抽出工程により抽出されたモチーフに基づいて、該モチーフを一部に含む遺伝子配列情報を、遺伝子配列情報データベースから検索する遺伝子配列情報検索工程と、

該遺伝子配列情報検索工程により検索された遺伝子配列情報を、前記遺伝子配列情報入力工程により入力された遺伝子配列情報に付け加える遺伝子配列情報付加工程と、

を含んで構成されたことを特徴とする遺伝子のモチーフ抽出処理方法。

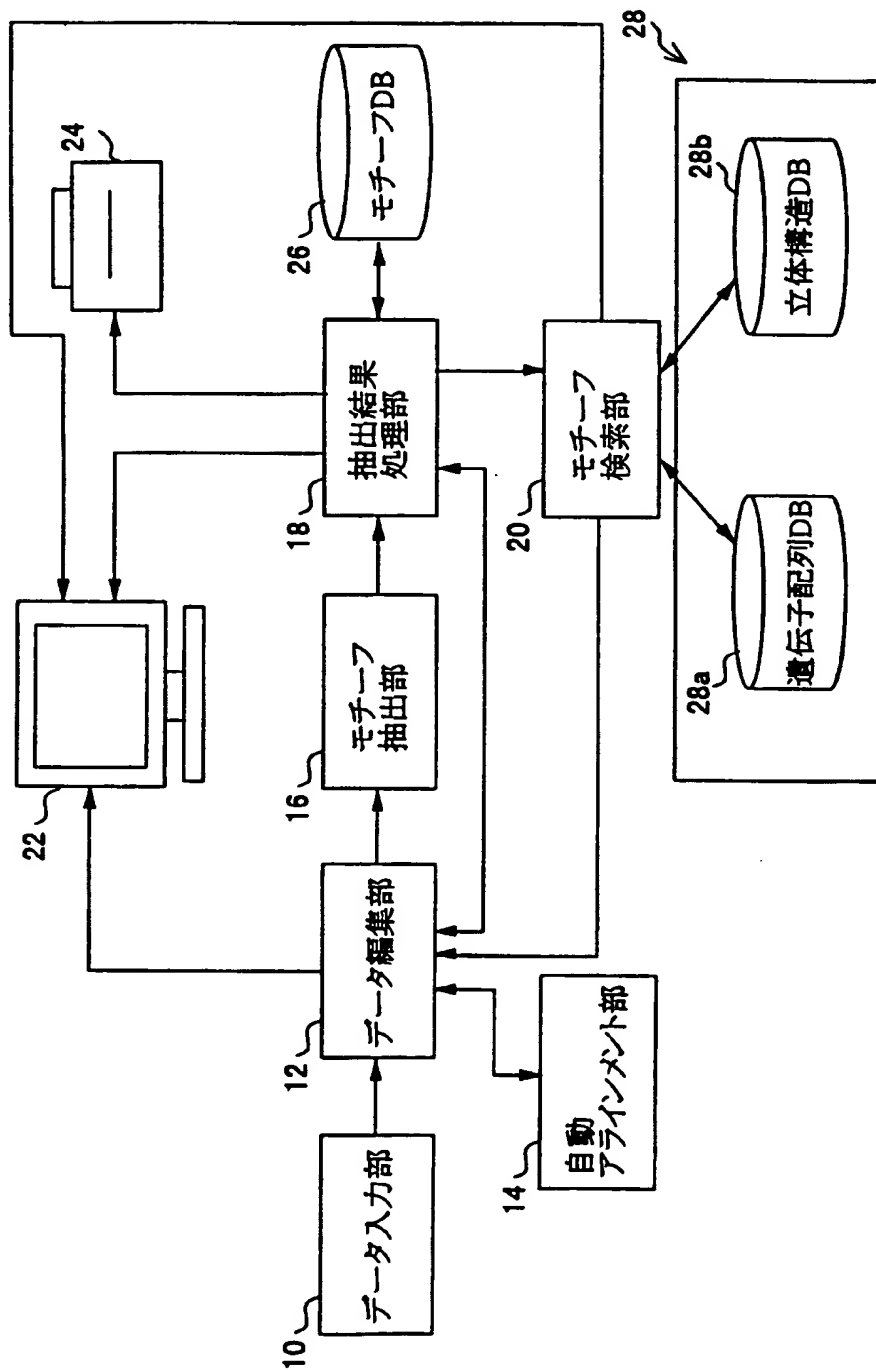
9. 少なくとも1つの遺伝子配列情報を入力する遺伝子配列情報入力機能と、  
該遺伝子配列情報入力機能により入力された遺伝子配列情報から、遺伝子のモチーフを抽出するモチーフ抽出機能と、

該モチーフ抽出機能により抽出されたモチーフに基づいて、該モチーフを一部に含む遺伝子配列情報を、遺伝子配列情報データベースから検索する遺伝子配列情報検索機能と、

該遺伝子配列情報検索機能により検索された遺伝子配列情報を、前記遺伝子配列情報入力機能により入力された遺伝子配列情報に付け加える遺伝子配列情報付加機能と、

を実現するための遺伝子のモチーフ抽出処理プログラムを記録した記録媒体。

図1



2/12

図2

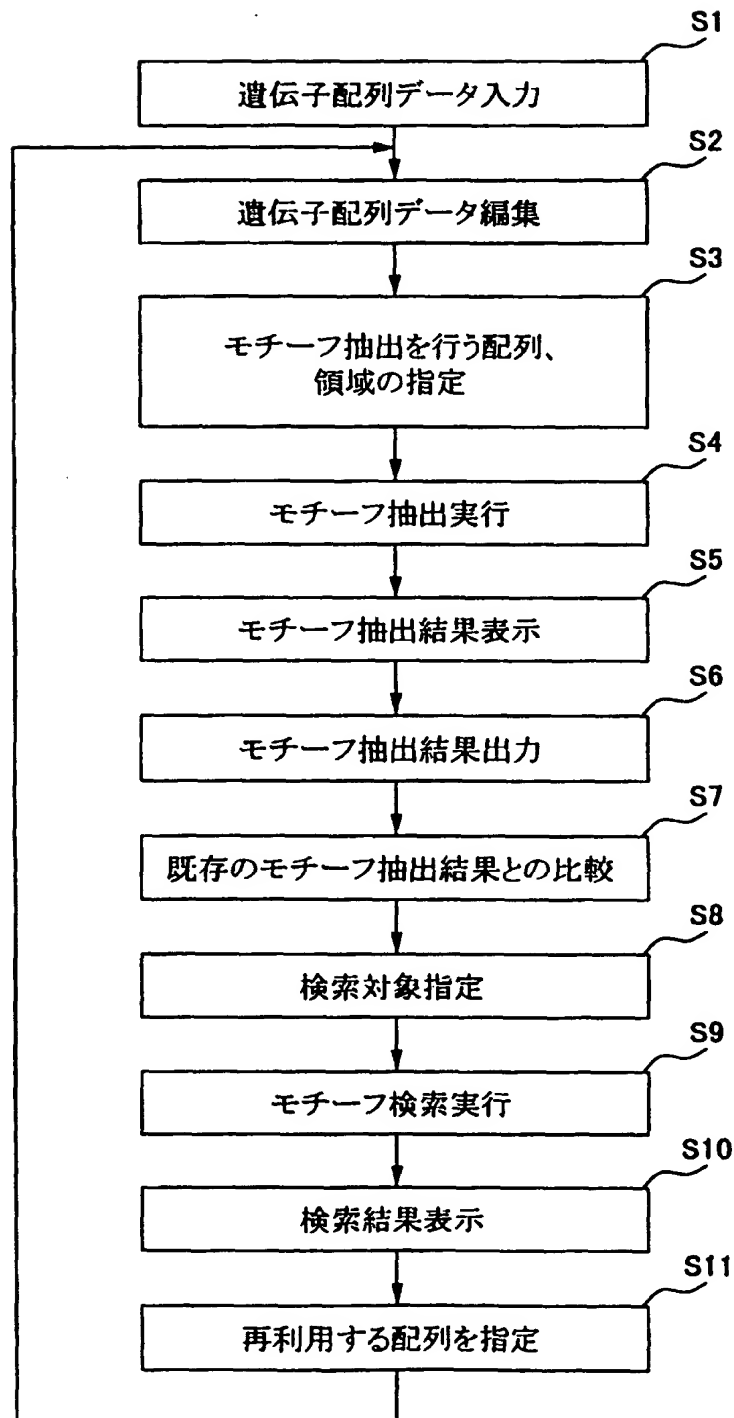


図3

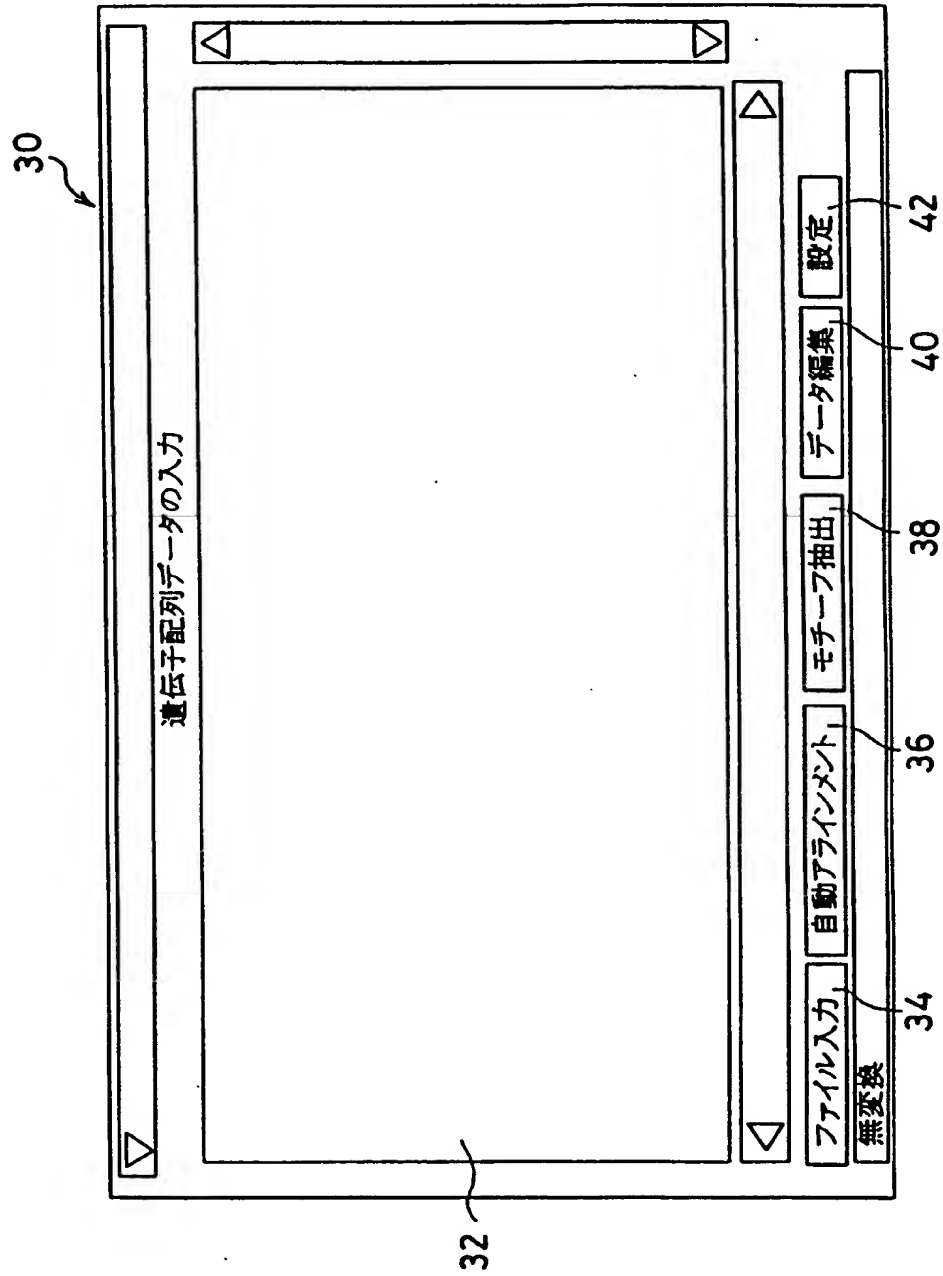


図4

30

遺伝子配列データの入力

XLN2MAB-1	10	20	30	40	50	60
BFU82487-1	.....	.....	.....	.....	.....	.....
BRENG1-1	.....	.....	.....	.....	.....	.....
BRENG2-1	.....	.....	.....	.....	.....	.....
BRENG20-1	.....	.....	.....	.....	.....	.....
BRENG3-1	.....	.....	.....	.....	.....	.....
MUSEN2AB-1	.....	.....	.....	.....	.....	.....
XLN2MAB-1	70	80	90	100	110	120
BFU82487-1	.....	.....	.....	.....	.....	.....
BRENG1-1	.....	.....	.....	.....	.....	.....
BRENG2-1	.....	.....	.....	.....	.....	.....
BRENG20-1	.....	.....	.....	.....	.....	.....
BRENG3-1	.....	.....	.....	.....	.....	.....
MUSEN2AB-1	.....	.....	.....	.....	.....	.....
XLN2MAB-1	130	140	150	160	170	180
BFU82487-1	.....	.....	.....	.....	.....	.....
BRENG1-1	.....	.....	.....	.....	.....	.....

32

ファイル入力

自動アライメント

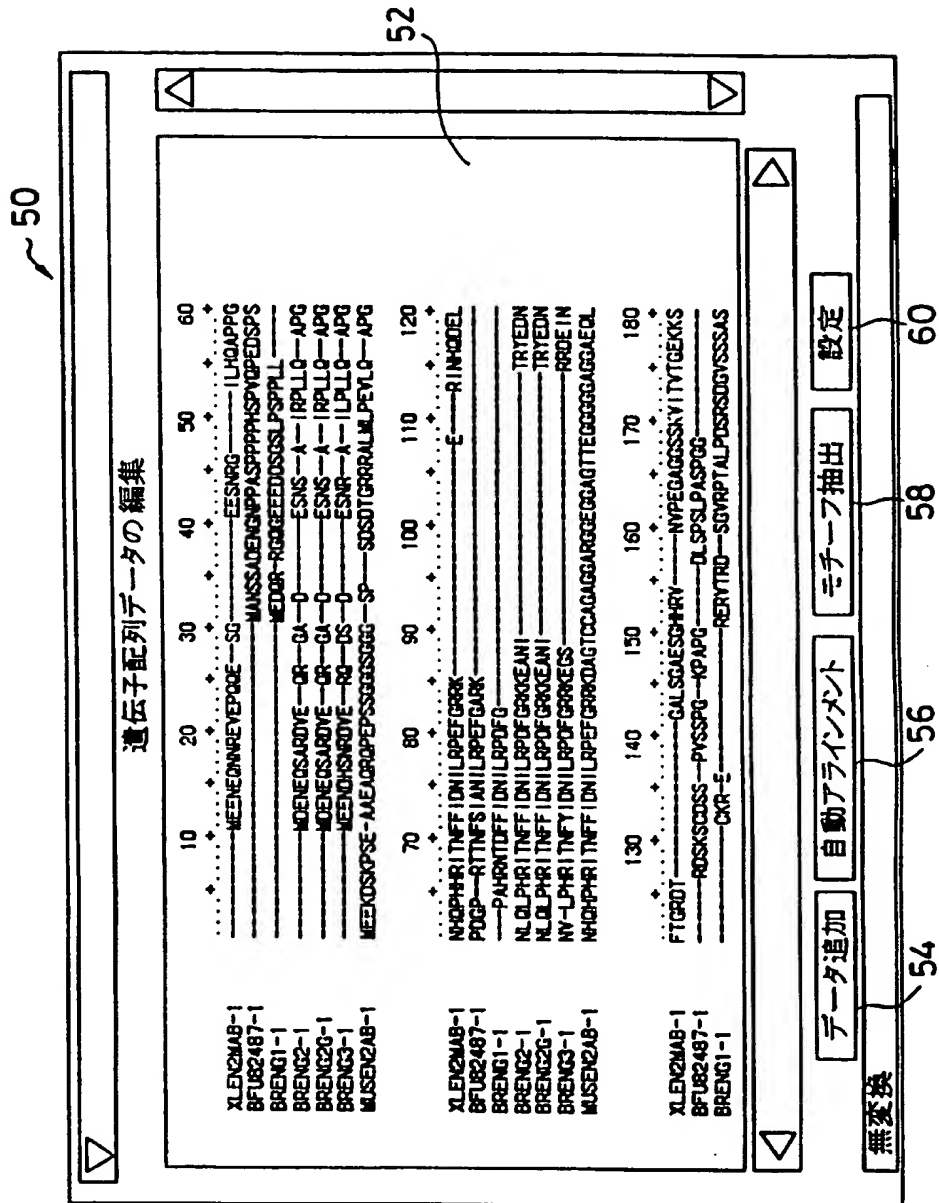
モチーフ抽出

データ編集

設定

無変換

図5





50

▽
遺伝子配列データの編集
▷

	10	20	30	40	50	60
XLEN2MAB-1	+	+	+	+	+	+
BFU82487-1	MEENE	NNRE	VEPQ	EE	NRG	---
BRENG1-1	---	---	---	---	---	---
BRENG2-1	MDNE	GSARD	VE	QR	GA	---
BRENG2G-1	MDNE	GSARD	VE	QR	GA	---
BRENG3-1	MEEN	HSNR	DE	RQ	DS	---
MUSEN2AB-1	MEEK	DSKP	SE	AAE	AD	---
	10	20	30	40	50	60
	+	+	+	+	+	+
XLEN2MAB-1	NHQPH	RITN	FIS	AN	LR	---
BFU82487-1	PDGP	---	---	---	---	---
BRENG1-1	---	---	---	---	---	---
BRENG2-1	NLQL	PHRI	TNFI	DNI	LR	---
BRENG2G-1	NLQL	PHRI	TNFI	DNI	LR	---
BRENG3-1	NV	---	---	---	---	---
MUSEN2AB-1	NHQPH	RITN	FIS	AN	LR	---
	10	20	30	40	50	60
	+	+	+	+	+	+
XLEN2MAB-1	FTGR	DT	---	---	---	---
BFU82487-1	---	---	---	---	---	---
BRENG1-1	---	---	---	---	---	---

◁

データ追加

自動アラインメント

モチーフ抽出

設定

▷

無変換

58



図7

50

遺伝子配列データの編集

XLEN2MAB-1  
BFU82487-1  
BRENG1-1  
BRENG2-1  
BRENG2G-1  
BRENG3-1  
MUSEN2AB-1

102030405060

MEENEONNREVEPQDE--SG--MANSSADENGNPPASPPPHSPVQPEDSPS  
-----EESNRG-----ILHQAPPG  
-----MEDQR--RGEGEEDDGSLSPPPL--  
-----MDENEGSARDVE--QR--GA--D--ESNS--A--IRPLLQ--APG  
-----MDENEGSARDVE--QR--GA--D--ESNR--A--IRPLLQ--APG  
-----MEENDHSNRDVE--RQ--DS--D--ESNR--A--IRPLLQ--APG  
MEEKDKPSE--AAEADQRQPEPSSGGGGG--SP--SDSDTGRRRALMLPEVLQ--APG

102030405060

NHQPHRI TNFFI DNI LRPEFGRRK  
PDGP--RTTNF SIANI LRPEFGARK  
-----PAHRNTDFI DNI LRPDG  
NLQLPHRI TNFFI DNI LRPDGRRKEANI  
NLQLPHRI TNFFI DNI LRPDGRRKEANI  
NV-LPHRI TNFYI DNI LRPDGRRKEG  
NHQPHRI TNFFI DNI LRPEFGRRKDGTCACAGGARGGEGGAGTTEGGGGGAGGAEQL  
-----FTGRDT-----GALSGAESGHHRV-----NVPEGAGGSSKVI TVTGEKKS  
-----RDSKSCDSS--PVSSPG--KPAPG-----DLSPSLPASPGG  
-----CKR-E-----RERVTRD--SGVRPTALPDSRSDGVSSSAS

102030405060

XLEN2MAB-1  
BFU82487-1  
BRENG1-1  
BRENG2-1  
BRENG2G-1  
BRENG3-1  
MUSEN2AB-1

102030405060

FTGRDT-----GALSGAESGHHRV-----NVPEGAGGSSKVI TVTGEKKS  
-----RDSKSCDSS--PVSSPG--KPAPG-----DLSPSLPASPGG  
-----CKR-E-----RERVTRD--SGVRPTALPDSRSDGVSSSAS

データ追加

自動アライメント

モチーフ抽出

設定

無変換

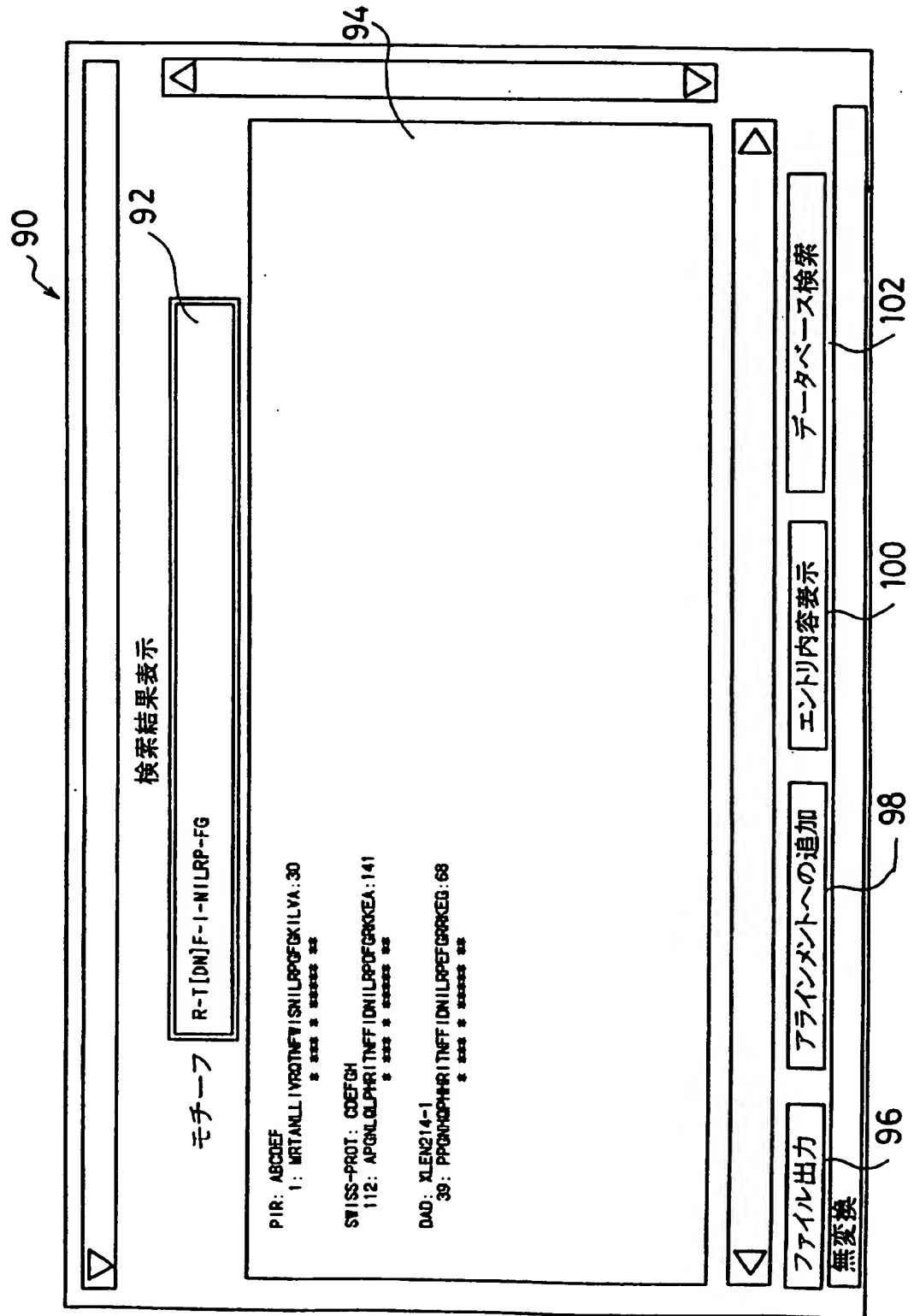
58





10/12

図10



11/12

図11

90

検索結果表示

モチーフ R-T(DN)F-I-NILRP-FG

PIR: ABCDEF  
1: MSTAALLIVGRTNFTISNLRPFQXILVA: 30  
\* 000 \* 000000 00

SWISS-PROT: COEFGH  
112: APQNLQLPRIITNFTIDNLRPFQKCEA: 141  
\* 000 \* 000000 00

x DAD: XLEN214-I  
39: PPGHGHPRITNFTIDNLRPFQKCEG: 68  
\* 000 \* 000000 00

98

ファイル出力 アライメントへの追加 エントリ内容表示 データベース検索

無変換

12/12

図12

50

### 遺伝子配列データの編集

	10	20	30	40	50	60
XLEN2MAB-1	.....	.....	.....	.....	.....	.....
BFUB2487-1	.....	.....	.....	.....	.....	.....
BRENG1-1	.....	.....	.....	.....	.....	.....
BRENG2-1	.....	.....	.....	.....	.....	.....
BRENG20-1	.....	.....	.....	.....	.....	.....
BRENG3-1	.....	.....	.....	.....	.....	.....
MUSEN2AB-1	.....	.....	.....	.....	.....	.....
XLEN214-1	.....	.....	.....	.....	.....	.....

	70	80	90	100	110	120
XLEN2MAB-1	.....	.....	.....	.....	.....	.....
BFUB2487-1	.....	.....	.....	.....	.....	.....
BRENG1-1	.....	.....	.....	.....	.....	.....
BRENG2-1	.....	.....	.....	.....	.....	.....
BRENG20-1	.....	.....	.....	.....	.....	.....
BRENG3-1	.....	.....	.....	.....	.....	.....
MUSEN2AB-1	.....	.....	.....	.....	.....	.....
XLEN214-1	.....	.....	.....	.....	.....	.....

	130	140	150	160	170	180
XLEN2MAB-1	.....	.....	.....	.....	.....	.....

データ追加
自動アラインメント
モチーフ抽出
設定

無変換

52

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/JP99/01693

A. CLASSIFICATION OF SUBJECT MATTER  
Int.Cl<sup>6</sup> C12N15/00, G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
Int.Cl<sup>6</sup> C12N15/00, G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP, 7-274965, A (Director General of National Institute of Genetics), 24 October, 1995 (24. 10. 95) & GB, 2283840, A & US, 5598350, A	1-9
A	JP, 10-187666, A (NEC Corp.), 21 July, 1998 (21. 07. 98) (Family: none)	1-9

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p>
--	---

Date of the actual completion of the international search  
29 June, 1999 (29. 06. 99)

Date of mailing of the international search report  
6 July, 1999 (06. 07. 99)

Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## 国際調査報告

国際出願番号 PCT/J P 99/01693

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>8</sup> C12N 15/00, G06F 17/30

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>8</sup> C12N 15/00, G06F 17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP, 7-274965, A (国立遺伝学研究所長) 24. 10月. 1995 (24. 10. 95) & GB, 2283840, A & US, 5598350, A	1-9
A	JP, 10-187666, A (日本電気株式会社) 21. 7月. 1998 (21. 07. 98) (ファミリーなし)	1-9

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」 同一パテントファミリー文献

国際調査を完了した日

29. 06. 99

国際調査報告の発送日

06.07.99

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

高堀 栄二



4 B 9281

電話番号 03-3581-1101 内線 3448